

Statistique-Mathématique
TP

Travail à faire en groupe de 3 personnes.
(remise sur Google Classroom)

Exercice 1 *Simulation de lois*

1. Simuler un échantillon de taille 10000 suivant une loi binomiale $\mathcal{B}(30, 0.5)$. Tracer l'histogramme de l'échantillon obtenu.
2. Simuler un échantillon de taille 10000 suivant une loi normale $\mathcal{N}(3, .9)$. Tracer la fonction de densité de l'échantillon obtenu. Choisir un intervalle contenant 0 pour domaine de représentation.
3. Simuler un échantillon de taille 10000 suivant une loi du χ^2 à 20 degrés de liberté. Tracer la fonction de densité de l'échantillon obtenu. Choisir un intervalle contenant 0 pour domaine de représentation.

Exercice 2 *Méthode de Monte Carlo*

Soit $I_2 = \int_0^1 \sqrt{1-x^2} dx$

Estimer I_2 par une méthode de Monte Carlo avec $n = 10000$

Observer par graphique l'évolution de cette estimation lorsque n varie et vérifier la cohérence avec la valeur théorique $I_2 = \frac{\pi}{4}$.

Exercice 3 Nous donnons les couples d'observations suivants :

x_i	18	7	14	31	21	5	11	16	26	29
y_i	55	17	36	85	62	18	33	41	63	87

1. La première étape est d'obtenir les données. Enregistrer-les dans un format adapté pour une lecture par la suite avec Python.
2. Représentez les y_i en fonction des x_i . A la vue de cette représentation, pouvons-nous soupçonner une liaison linéaire entre ces deux variables?
3. Déterminer pour ces observations la droite des moindres carrés, c'est-à-dire donner les coefficients de la droite des moindres carrés.
4. Donner les ordonnées des y_i calculés par la droite des moindres carrés correspondant aux différentes valeurs des x_i .
5. Tracer ensuite la droite sur le même graphique.
6. Quelle est une estimation plausible de Y à $x_i = 21$?
7. Quel est l'écart entre la valeur observée de Y à $x_i = 21$ et la valeur estimée avec la droite des moindres carrés? Comment appelons-nous cet écart?

8. Est-ce que la droite des moindres carrés obtenue en 2. passe par le point (\bar{x}, \bar{y}) ? Pouvons-nous généraliser cette conclusion à n'importe laquelle droite de régression?

Exercice 4 (Données covid-19 Sénégal)

Dans un travail, nous utiliserons les données qui se trouve dans le fichier "regions_cas.csv", qui contient le nombre de malades de covid-19 par région entre le 29 mars 2020 et le 20 Octobre 2020.

1. Lire les données dans Python. On créera un dataframe contenant les variables suivantes : date, region et malades. On supprimera les accents et les ".". En suite seule la première lettre du mot sera en majuscule.(par exemple on remplace THIÈS par Thies, KÉDOUGOU par Kédougou).
2. Convertir la variable date en type datetime, et supprimer toutes les lignes ayant des valeurs manquantes (s'il en existe).
3. Créer une fonction qui retourne un dataframe à 3 colonnes (date, region, maladesparegion). La dernière colonne contiendra le nombre de malades de covid-19 par régions aux différentes dates données.
4. Supposons que la variable *maladesparegion* suit une loi de Poisson de paramètre λ . Estimer λ .
5. Utiliser un test statistique pour vérifier si la variable *maladesparegion* suit une loi de Poisson.
6. Supposons que la variable *maladesparegion* suit une loi de Binomiale négative de paramètres r et p . Estimer r et p .
7. Créer une fonction *CarteRegions(madate)* qui affiche la carte choroplèthe des régions en utilisant le nombre de malades.(Vous pouvez télécharger les coordonnées géographiques du Sénégal dans le site <https://gadm.org/data.html>)